Not your grandparents' confidence intervals

Gregor Reich (NHH) and Ken Judd (Hoover)

March 23, 2021

How can we determine the statistical properties of our estimators?

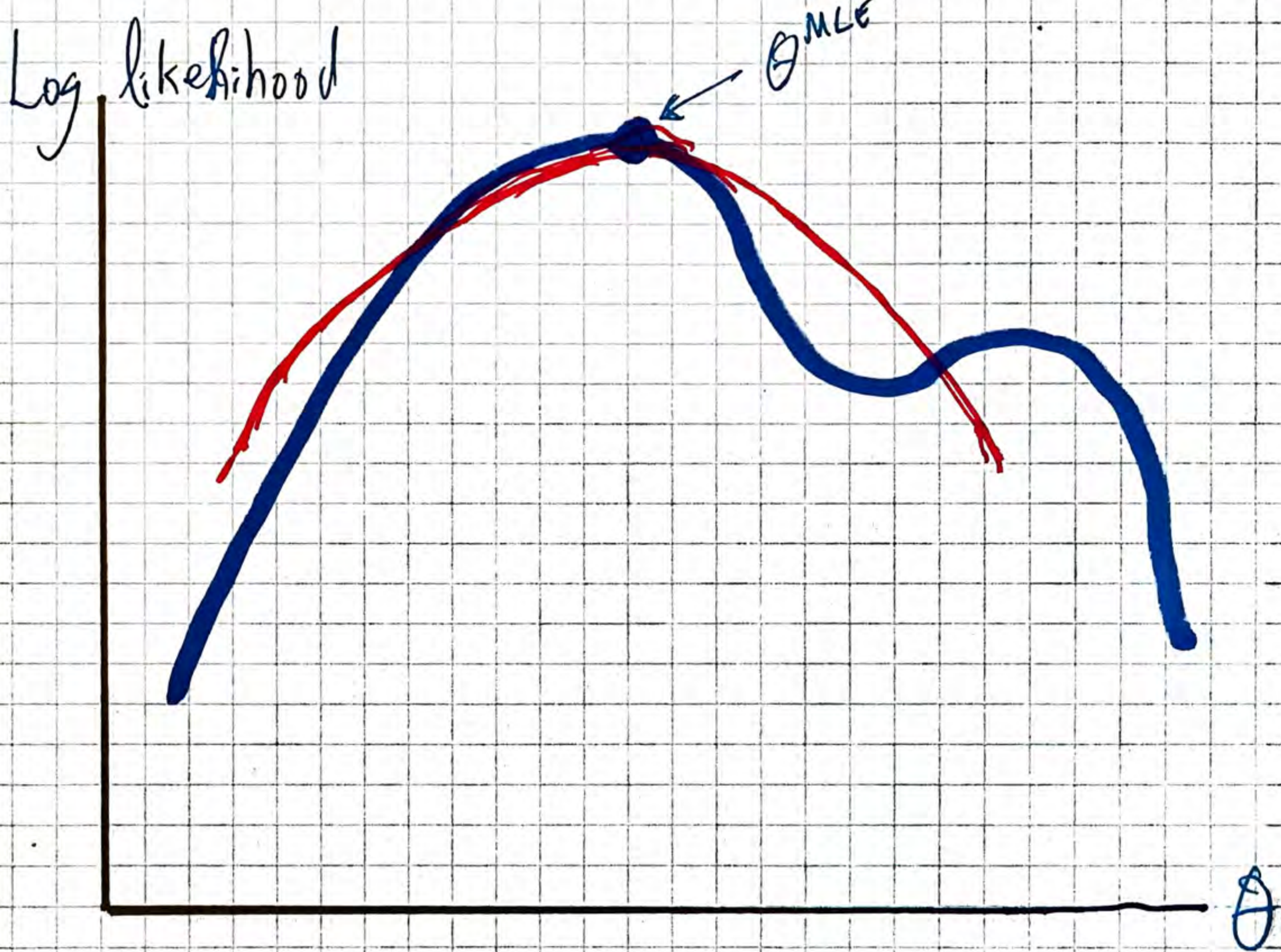
- Repeat the experiment many times
  - Pick true parameters
  - Generate synthetic data sets of various sizes
  - Apply procedures
  - Record results and fit to some class of distributions
  - Who needs theoretical econometricians?
- Problem: we aren't allowed to get the required computer power
  - We didn't build the bomb
  - Current users do not want new users
- Econometricians to the rescue
  - They develop theories
  - They use asymptotic properties to derive useful statistical tests

Problem 1

- Asymptotically, we are all dead

Problem 2

- We have finite sample problems during our finite life



Log likelihood
θMLE
θ

Today

- Review basic statistics to remind ourselves of the subtle differences in concepts
- Describe Reich-Judd approach which avoids some of the approximations typically used
- Describe application to ... what else .... Zurcher bus model

Statistics and estimates

- Let $f(x;\theta)$ denote a family of probability masses or density functions over S – potentially multivariate – parameterized by θ. Suppose the random variables $X^1, \ldots, X^n$ are independently and identically distributed according to some $f(x;\theta_0)$ for some θ_0. Let $X^{1:n} = (X^1, \ldots, X^n)$ denote a collection of random variables for some n. The data matrix $x^{1:n} = (x^1, \ldots, x^n)$ is called a *realization* of the *random sample of size n*
- Consider a real function $h(\cdot)$; the random variable $T_n = h(X^{1:n})$ with realization $t_n = h(x^{1:n})$ is called a *statistic*. If a sequence of statistics, T_n, is used to infer an unknown parameter θ, it is called an *estimator*; when appropriate, it can be denoted by $\hat{\theta} = \hat{\theta}(X^{1:n})$. A concrete value for such an estimator based on $x^{1:n}$ is called an *estimate*, either denoted by t or θ.

Standard error and confidence interval

- Let T_n be an estimator of θ, and V be a consistent estimator of its variance $\text{var}(T_n)$. Then, the *standard error*, $(T) \equiv \sqrt{V}$, is a consistent estimator of its standard deviation $\sqrt{\text{var}(T)}$.
- Given a fixed $\gamma \in (0,1)$, the two statistics $T_{n,l}$ and $T_{n,b}$ form the boundaries of a $\gamma \cdot 100\%$ *confidence interval* if $P(T_{n,l} \leq \theta \leq T_{n,l}) = \gamma \; \forall \theta \in \Theta$; γ is called the *confidence level*, or alternatively the *coverage probability*.
- Comments
  - The main difficulty with standard errors is obtaining a consistent estimator V of the variance of the estimator T_n
  - Finding a statistic that fulfills the coverage condition is generally nontrivial. Most of the time, general statistics that rely on asymptotics will be used.
  - The correct interpretation of a confidence interval is that if the random sampling in the population were to be repeated, $\gamma \cdot 100\%$ of the confidence intervals obtained would cover the true parameter θ.
  - It is *not* correct to say that given a sample, the confidence interval contains the true parameter with $\gamma \cdot 100\%$ probability, as there is no randomness involved anymore once the sample is taken.

z-Statistic, Asymptotic Normality, Wald Confidence Interval

- If T_n is a consistent estimator for θ, the z-statistic will—under appropriate regularity conditions—by asymptotically (standard) normal distributed:

$$Z(\theta) \equiv \frac{T_n - \theta}{(T_n)}(0, 1).$$

- The two statistics $T_n \pm z_{\frac{1+\gamma}{2}}$ form the boundary of an approximate $\gamma \cdot 100\%$ confidence interval, also referred to as the Wald confidence interval, where z is the corresponding quantile of the standard normal distribution.

Likelihood function

- The *likelihood (function)* is the joint probability or the joint density of the data, given a particular value of the parameter, written as a function of the parameter (fixing the data): $L(\theta; x) \equiv f(x; \theta)$
- The *maximum likelihood estimate* is defined as $\theta^{ML} = \arg\max_{\theta \in \Theta} L(\theta; x)$, and the *maximum likelihood estimator* as $\hat{\theta}^{ML} = \arg\max_{\theta \in \Theta} L(\theta; X)$; both objects might be abbreviated by MLE.
- Due to the independence of the draws, the likelihood function for the sample is the product of the individual likelihoods: $L(\theta; x^{1:n}) = \prod_1^n f(x^i; \theta)$.
- Every monotone transformation of L has the same extremal values
  - We often use the natural logarithm of the likelihood, called the *log-likelihood*: $l(\theta; x) \equiv \log(L(\theta; x))$.
  - Since the log of a product is a sum, maximizing the log-likelihood avoids problems of underflow.

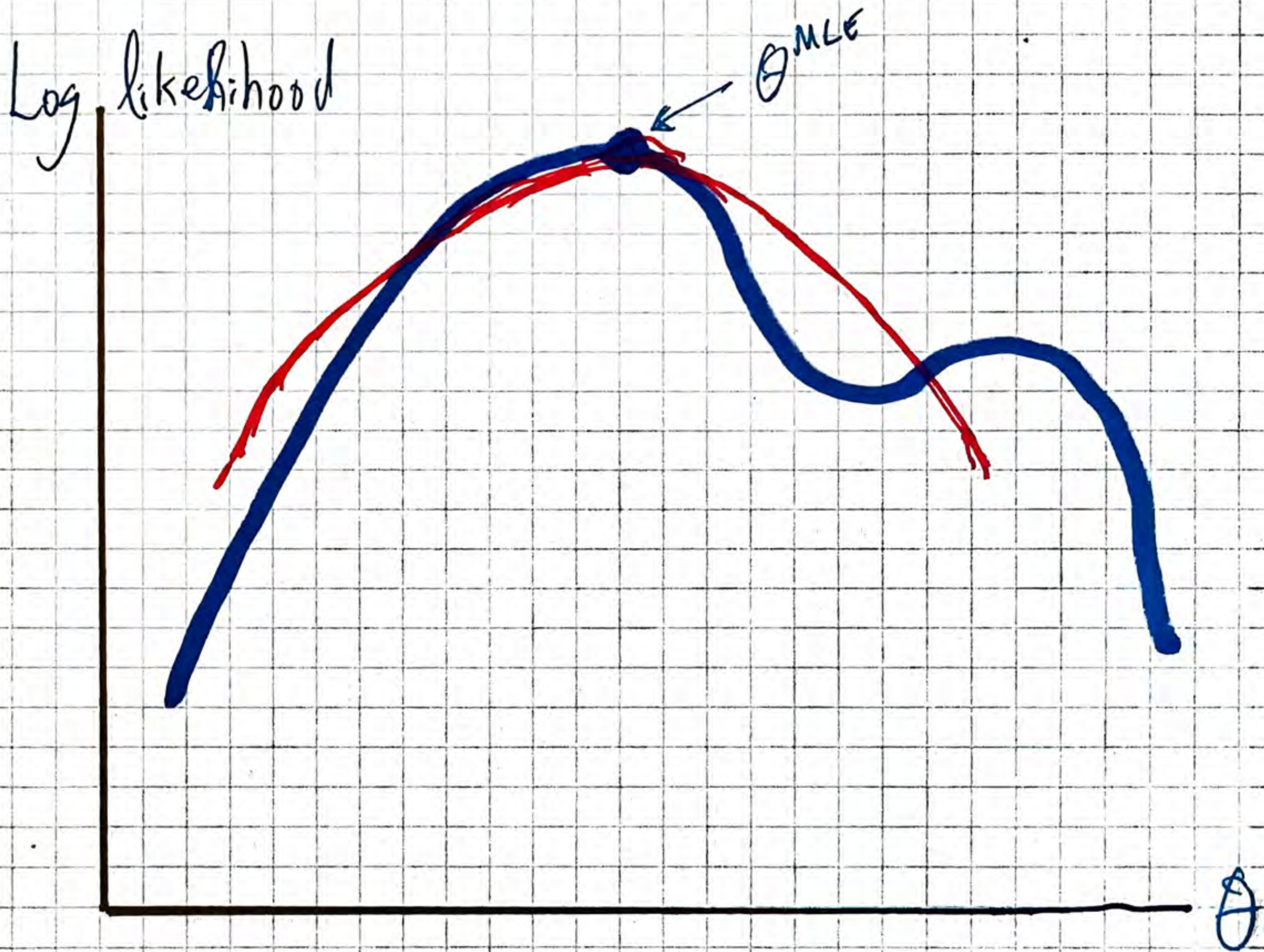
Relative likelihood

- The relative likelihood is defined by $\tilde{L}(\theta; x) = \frac{L(\theta;x)}{L(\hat{\theta}_{ML};x)}$. In particular, $0 \leq \tilde{L}(\theta; x) \leq 1$.
- The following definitions give names to the first and second derivatives of the log-likelihood function:
  - *Score function*: $S(\theta; X) \equiv \frac{dl(\theta;x)}{d\theta}$
  - *(Ordinary) Fisher information*: $I(\theta; X) \equiv -\frac{d^2 l(\theta;x)}{d\theta^2} = -\frac{dS(\theta;X)}{d\theta}$
  - *Expected Fisher information*: $\mathbb{E}(I(\theta_0; X))$, where the expectation is taken with respect to X (this implies that the expectation is integrated with against $f(X, \theta_0)$ at the true parameter value).
  - *Observed Fisher information*: $I(\hat{\theta}^{ML}; X^{1:n})$ (at the ML estimator)
- Asymptotic Normality of ML Estimator Suppose $\hat{\theta}^{ML}$ is a consistent estimator for the true parameter θ_0, and the Fisher regularity conditions hold. Then,

$$\sqrt{n \cdot J(\theta_0)}(\hat{\theta}^{ML} - \theta_0)(0, 1)$$

Wald statistics and confidence interval

- To test $H_0 : \theta_0 = \tilde{\theta}_0$, the *Wald statistic* is defined by $\sqrt{I(\hat{\theta}^{ML}; X^{1:n})}(\hat{\theta}^{ML} - \tilde{\theta}_0)$, which is asymptotically (standard) normal distributed.
- The bounds of the $\gamma \cdot 100\%$ *Wald confidence interval* are obtained as $\hat{\theta}^{ML} \pm z_{\frac{1+\gamma}{2}}(\hat{\theta}^{ML})$
- The Wald confidence interval is generally considered to be "too large" for a given γ.
- It is not invariant to non-linear transformations because the Wald statistic is based on a second order approximation of likelihood, and does not involve the likelihood function itself



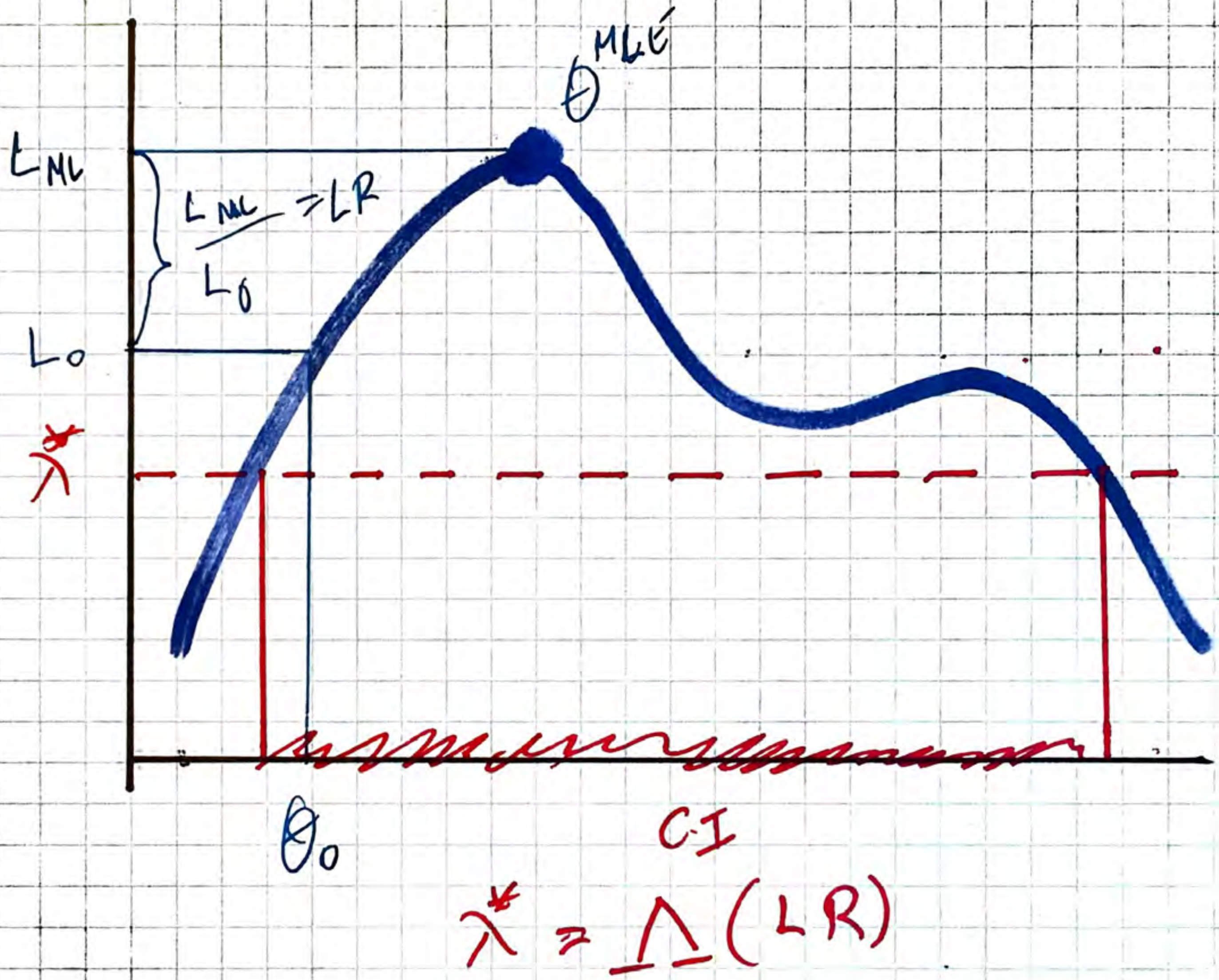
Log likelihood
θMLE
θ

Likelihood ratio statistics

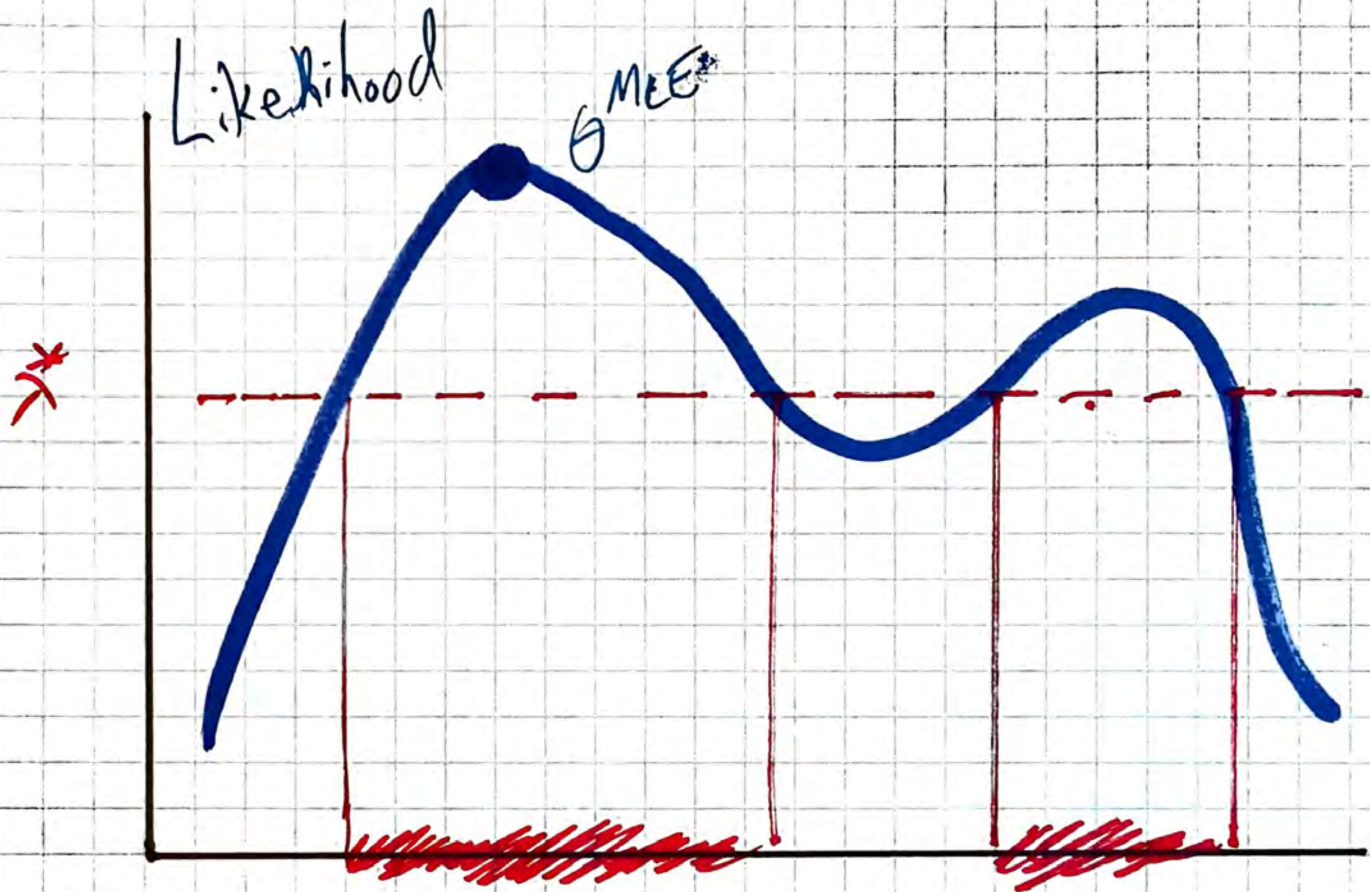
- The likelihood ratio statistic asymptotically follows a Chi-squared distribution with one degree of freedom:

$$-2(l(\hat{\theta}^{ML}; X) - l(\theta_0; X)) \equiv -2\tilde{l}(\theta_0; X)\chi^2(1).$$

- The set $\left\{\theta : \tilde{l}(\theta; X) \geq -0.5\chi^2_\gamma(1)\right\}$ forms the $\gamma \cdot 100\%$ *likelihood ratio (LR) confidence interval* for θ, where $\chi^2_\gamma(1)$ is the corresponding quantile of the Chi-squared distribution with one degree of freedom.
- The likelihood ratio confidence interval defines a *manifold*.
  - Numerical methods are required to approximate its *boundary*
  - In one dimension, finding the boundary boils down to finding an even number (usually 2) of solutions to a one dimensional equation.
  - Computing these confidence intervals as the solution to the likelihood ratio statistic equaling a quantile of the Chi-squared distribution is also referred to as *test inversion*, because one seeks the one value of the likelihood ratio such that the inequality holds strictly.
- *Wilk's theorem* generalizes this to multiple dimensions



θ MLE
L ML
L ML / L 0 = LR
L 0
λ*
θ0
C.I
λ* = Λ(LR)



Likelihood
θ^{MLE}
λ^*

Reich and Judd idea

- Let's compute the likelihood level set of the Chi-squared quantile
- Yes, let's compute manifolds
- Today we stay with one dimensional manifolds
- Perhaps you will see the multidimensional version next year